

# MPEG-21 DIA 기반 관심영역 추출을 위한 비디오 적응 엔진의 구현

\*박성준, \*손유미, \*Hendry, \*정현, \*Qonita, \*김종남, \*박근수, \*김문철

\* 한국정보통신대학교 멀티미디어 컴퓨팅, 통신 및 방송 연구실

\* KBS 기술연구소

## Implementation of MPEG-21 DIA based Video Transcoding Engine for ROI extraction

\*Seongjoon Pak, \*Yumi Sohn, \*Hendry, \*Hyun Jeong, \*Qonita, \*Jong-Nam Kim, \*Keunsoo Park and \*Munchurl Kim

\*Multimedia Computing, Communication and Broadcasting Laboratory, Information and Communications University (ICU)

\*Technical Research Institute, Korean Broadcasting System (KBS)

E-mail : rinusia@icu.ac.kr, Phone : 042-866-6246

### 요 약

MPEG-21은 사용자가 다양한 멀티미디어 데이터를 네트워크상에서 생성, 변형, 전달, 소비하기 위한 통합적 멀티미디어 프레임워크를 정의한 국제 표준이다. MPEG-21 DIA(digital item adaptation)는 멀티미디어 데이터를 사용자의 환경, 단말기 특성, 네트워크 특성에 대한 정보를 기술할 수 있는 표준을 정의하고 있다. 본 논문에서는 DIA에서 정의한 여러 가지 적응 변환 중 사용자의 관심 영역을 중심으로 한 의미론적 변환 엔진을 구현하였으며, 이에 대한 실험결과를 제시한다. 기존의 변환 방법이 비디오 프레임의 전체 크기, 해상도 화질을 조정하였다면, 의미론적 변환은 사용자의 관심 영역을 중심으로 변환을 수행 함으로써, 사용자의 관심도를 반영할 수 있는 장점이 있다.

### 1. 서 론

과거의 멀티미디어 콘텐츠 소비는 주로 고정된 환경에서 이루어져 왔다. 네트워크 및 단말기의 꾸준한 발전은 사용자가 다양한 종류의 멀티미디어 콘텐츠를 다양한 환경에서 소비할 수 있도록 하였다. 다양한 네트워크 상황 및 사용자 환경은 사용자가 콘텐츠를 소비하는데, 여러가지 운영체제 및 코덱, 단말기, 네트워크 종류 및 속도 등을 고려하게 만들었으며, 범용적 멀티미디어 접근을 어렵게 하였다. 이에 MPEG-21은 다양한 멀티미디어 콘텐츠를 다양한 환경에서 사용자가 상호 호환적인 방법으로 쉽고 편리하게 생성, 배급, 소비할 수 있도록 하는 방법인 '멀티미디어 프레임워크'를 정의 하여 이러한 문제에 해답을 제시하고자 하였다.

MPEG-21의 DIA(digital item adaptation)에서는 상호 호환성이 높은 투명한 접속을 가능하게 하기 위해 필요한 다양한 적응 변환 방법을 정의하고 이를 위한 부가 정보기술 방법을 기술하고 있다. 적응 변환은 기술정보 적응변환과 자원(資源) 적응변환으로 이루어 진다. 기술 정보는 멀티미디어 콘텐츠를 적응변환 하기 위한 각종 정보를 담은 메타데이터 정보이며, 자원은 멀티미디어 데이터 그 자체이다. 적응변환 엔진의 설계는 MPEG의

표준화의 대상이 아니다. 다만 DIA에서는 적응변환을 위한 메타데이터 기술 방법을 표준화 하여, 표준을 기반으로 제작한 어떠한 적응변환도 서로 호환 되도록 하고 있다. 본 논문에서는 실제의 콘텐츠 소비과정에서 적용 가능한, 사용자 관심 영역에 따른 트랜스코딩 엔진을 제안한다.

전통적인 트랜스코딩 방법은 SNR을 줄이거나 콘텐츠의 시간 또는 공간 축을 변화시켜 양질의 자원을 낮은 질로 변화시키는 것이다. B-picture을 생략하고 I-picture만 사용하는 B-프레임 제거법(frame dropping)이나 frame의 x, y 픽셀 수를 조정하는 것이 그러한 방법에 해당된다. 실제로 인터넷방송을 볼 때 네트워크 상황에 따라서 그림이 끊기는 현상이 나타나는데 이러한 방법이 시간에 따른 트랜스코딩(temporal scalability)에 해당된다. 이에 반해서, 사용자 관심영역 기반의 비디오 트랜스코딩은 사용자의 관심도에 따라서 임의의 영역을 확대해서 보여주는 방식이다. 예를 들어 축구장 전경을 내용으로 하는 영상을 사용자 단말의 크기에 따라 비례적으로 공간적 해상도를 축소할 경우, 축구장 내의 움직이는 선수들의 크기 또한 매우 작게 되어 영상의 내용적 의미가 이해되기 어렵게 될 수도 있다. MPEG-21 DIA 표준안에서는 멀티미디어 콘텐츠의 의미론적 트랜스코딩 기술을 위해서 *PresentationPreferenceDS*에 *FocusOfAttention* 이란 요소가 포함되어 의미론적 트

랜스 코딩의 기술정보가 상호 유통 가능하도록 하고 있다.

이러한 관심영역에 따른 적응변환 엔진(Region Of Interest adaptation engine)을 설계할 고려할 점은 다양한 비디오 코덱 때문에 한 가지 종류의 미디어형만 지원하는 적응변환 엔진은 실효성이 떨어진다는 것이다. 본 논문에서는 다양한 미디어 형을 지원하는 범용적 사용자 관심영역 기반의 적응변환 엔진을 제안한다. 본 논문은 다음과 같이 구성되어 있다. 우선 2장에서는 관심영역 기반의 적응변환 엔진의 기본 개념에서 대해서 설명을 하며, 요소 기술인 MPEG-21 DIA와 사용자 관심영역의 기술방법에 대해서 설명한다. 3장에서는 사용자 관심영역 기반의 적응변환 엔진을 설계하고 이를 설명한다. 마지막으로 4장에서는 제안된 ROI 엔진으로 제작 가능한 어플리케이션을 구현하여 이에 대한 실험결과를 제시한다.

## II. ROI에 기반한 의미론적 적응변환

### MPEG-21 DIA(Digital Item Adaptation)

MPEG-21 파트7 규격인 MPEG-21 디지털 아이템 적응(DIA: Digital Item Adaptation)은 디지털 아이템을 사용자 특성과 환경 정보, 네트워크나 터미널의 특성을 고려하여 다양한 멀티미디어 콘텐츠의 소비를 가능하게 하기 위하여 만들어졌다. DIA 처리 과정은 기능적으로 크게 '리소스 적응 엔진(resource adaptation engine)'과 '서술 메타데이터 적응 엔진(description adaptation engine)'으로 분류 될 수 있는데, 사용자 특성과 사용자 터미널 및 네트워크 환경 등의 사용자 환경 정보를 기술한 'DIA 서술자 툴(DIA description tool)'에 기반하여 입력된 디지털 아이템은 리소스 및 서술자 변환 과정을 거쳐 적응된(adapted) 디지털 아이템으로 출력된다.

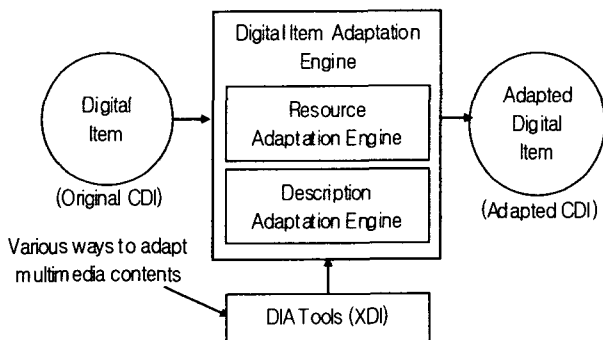


그림 1 디지털 아이템 적응 변환의 개념

그림 1은 MPEG-21 디지털 아이템 적응 변환의 개념도이다. 리소스적응엔진과 메타데이터 적응엔진이 하나의 DIA 적응변환 엔진을 구성함을 나타낸다.

본 논문에서는 영상 데이터의 적응변환을 위해서, DIA

에서 정의한 적응변환의 한 부분인 사용자 관심 영역에 따른 적응변환 엔진을 설계한다. 이를 위해서 우선 다음 장에서는 ROI 정보 표현방법에 대해서 설명한다.

### 비디오 데이터에서의 사용자 관심영역의 표현

지금까지 여러가지 비디오 데이터를 트랜스 코딩하는 방법이 존재 해왔다. 비디오 데이터가 대역폭이 가변적인 네트워크로 전달되기 위해서는 데이터의 품질을 조절함으로써, 정보량을 변동시키는 기능은 필수적이다. 사용자 단말이 사용자 터미널의 디스플레이 사이즈, 컴퓨팅 자원, 네트워크 대역폭 등으로 제한된 상황일 경우, 원래의 멀티미디어 콘텐츠가 그대로 사용자 단말에 제공되어도 제한된 단말기 상황으로 인하여 사용자는 원본 콘텐츠의 품질을 느낄 수 없다. 예를 들어, 축구 경기의 전경을 그대로 크기를 작게 해서 작은 디스플레이로 보여주는 경우 선수들이 작은 점들로 바뀌어서 시각적으로 제한된 영상만을 볼 수 있다. 이러한 경우에는 원본 콘텐츠를 그대로 리사이징 하기 보다, 사용자의 선호 영역을 중심으로 관심영역만 보여줄 필요가 있다.

사용자의 관심 영역은 그림 2에서 보는 바와 같이 MPEG-21 DIA 스키마의 *Presentation PreferenceDS* 의 *FocusOfAttention*이란 요소를 사용하여 기술한다. *FocusOfAttention* 의 하위 요소로 정지 관심 영역의 표현을 위한 *FocusOnStillRegion*과 이동 관심 영역의 표현을 위한 *FocusOfMovingRegion* 요소가 있다. 이들 요소는 비디오 데이터에서 관심 영역을 상자형태의 좌표로 표현한다.

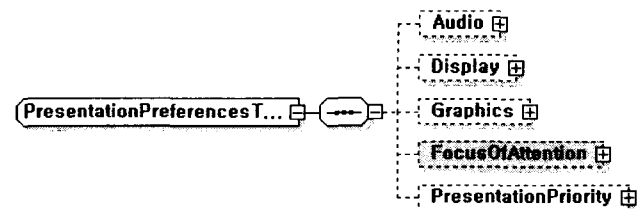


그림 2 PresentationPreference DS 의 구조

```
<?xml version="1.0" encoding="UTF-8"?>
<Mpeg7 xmlns="um:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="um:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
<DescriptionUnit xsi:type="StillRegionType"
id="original/baseball.mpg">
<SpatialMask>
<SubRegion><!-- just pre-define each duration gap -->
<Box>120 140 328 300</Box>
</SubRegion>
</SpatialMask>
<MediaRelIncrTimePoint
```

```
mediaTimeUnit="PT 1N30F">561</MediaRelIncrTimePoint>
</DescriptionUnit>
</Mpeg7>
```

그림 3 사용자 관심영역의 메타데이터 표현 예제

그림 3은 사용자 관심 영역 정보를 MPEG-7의 스키마에 따라서 표현한 예제이다. MPEG-21 DIA FocusOfAttention 요소에서는 사용자 관심영역의 적응 변환을 위하여, MPEG-7의 사용자 관심 영역 정보 스키마를 참조하고 있다. MediaRelIncrTimePoint 요소는 미디어 데이터에서 적응변환의 시작점을 초단위로 정의하는 부분이다. MediaRelIncrTimePoint 를 통하여 2시간 영화에서 사용자 관심 영역의 적응변환이 15분째 frame부터 시작되도록 할 수 있다.

### III. 제안된 사용자 관심영역에 따른 적응변환 시스템

사용자 관심 영역에 따른 적응변환 엔진을 설계하는데 가장 큰 어려운 점은 다양한 멀티미디어 데이터를 수용해야 하고, 구조에 있어서 유연해야 한다는 것이다. 예를 들어 관심 영역변환이 MPEG-2를 표현하는 엔진처럼 그 기능이 특정 데이터 형의 특정 역할에 한정된다면, MPEG-21 DIA의 기본 개념인 다양한 멀티미디어의 상호호환적이고 투명한 소비라는 취지에 부합하지 못하는 구현이 된다.

멀티미디어 데이터를 이용해서 적응변환 엔진을 설계하는데 고려해야 하는 요소는 아래와 같다.

- 신속한 데이터 처리 - 멀티미디어 데이터는 대부분 대용량이다. 이에 대한 실시간 변환이 가능해야 한다.
- 정확한 동기 - 음성 및 비디오, 부가 데이터 간에 동기가 완벽해야 한다.
- 다양한 입출력 - 멀티미디어 데이터는 다양한 소스로부터 입력 가능하고 다양한 방법으로 출력이 가능해야 한다.
- 다양한 포맷 - 다양한 포맷(MPEG, AVI, DVD, WAV, MP3)을 처리할 수 있어야 한다.(일부 코덱은 계속 새로 등장하고 있다)

다양한 멀티미디어 데이터를 지원하기 위해서, 사용자 관심 영역 적응 변환 엔진은 그림 4 와 같은 구조로 설계되었다.

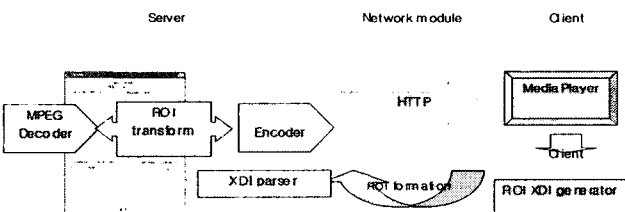


그림 4 전체적인 어플리케이션의 구조도

사용자는 클라이언트에서 원래의 콘텐츠를 스트리밍을 통하여 감상한다. 만일 사용자가 ROI 적응변환된 콘텐츠를 소비하길 원하면 ROI 영역을 지정해서 서버에 알려준다. 이때 ROI 정보는 그림 3과 같이 MPEG-21 DIA 표준에 따라서 기록한다. 서버 측에서는 사용자가 보낸 XDI를 파싱하여 ROI정보를 추출한다. 추출된 ROI 정보는 ROI 적응변환엔진에게 전달되고, ROI 적응변환된 비디오를 생성함과 동시에 스트리밍을 통해서 클라이언트에게 전달된다.

ROI 적응 변환엔진은 소프트웨어적으로 컴퍼넌트 구조를 이용하여 다양한 인코더 및 디코더 결합이 용이하도록 되어 있다. 또한 폭넓게 쓰이는 영상 데이터 형인 YUV 영역에서 적응변환을 수행함으로 유연성을 높였다. ROI 정보는 MPEG-21 DIA 표준에 따라서 기술된다. 표준을 따르는 단말기 사이에서 사용자 관심 영역정보가 자유로이 교환될 수 있다.

본 구현에는 사용자 관심영역에 따라 비디오 데이터를 적응 변환하는 ROI 변환 엔진과 더불어서, 사용자 (클라이언트)와 ROI 정보를 클라이언트와 주고 받기 위한 네트워크 모듈, XDI(context digital item) 파일을 파싱하기 위한 파싱 모듈 및 변환된 비디오 데이터를 스트리밍하기 위한 스트리머를 가지고 있다.

제안된 ROI 변환 엔진의 세부 구조는 아래 그림 5에 나타내었다.



그림 5 제안된 ROI 필터와 적응변환 된 콘텐츠를 얻기 위한 필터 배열

그림 5와 같이 비디오 디코더 뒤에 ROI 변환엔진이 연결되어 XDI에 따라 사용자 관심영역을 뽑아내고 이후 비디오 랜더링 필터를 통해서 이를 화면에 표시한다.

본 구현은 MPEG-21 DIA의 관점에서 보면, 사용자 관심영역 적응변환 엔진 중 resource 변환 엔진에 해당된다.

기술정보 변환 엔진은 MPEG-21 DIA에 따라서 기술된 ROI 정보를 읽어서 파싱해 주는 부분으로서 MS XML 파서를 이용하여 구현되었다.

ASF 파일은 윈도우 스트리밍 포맷으로서 동시 쓰기 및 읽기를 지원한다. 그러므로 손쉽게 HTTP를 이용한 스트리밍 시스템을 꾸밀 수 있다. 클라이언트 어플리케이션에서 사용자가 ROI정보를 생성하였을 때 이를 서버로 전송하면, 서버에서는 ROI 정보에 따라서 적응 변환된 스트리밍 영상을 클라이언트로 스트리밍 할 수 있다.

대개의 비디오 디코더는 다양한 형태의 출력 데이터 형을 제공한다. 예를 들면, RGB32, RGB565, YV12, YUV 등이 가능하다. 이러한 데이터 형은 대부분, RGB, YUV 데이터의 변형들이다. 본 구현에서는 YV12를 지원하도록 ROI변

환엔진을 구현하였다. YV12는 8비트 Y 데이터와 그림 크기가 1/2인 8비트의 U, V 데이터로 구성된다. 비디오 디코더의 출력형식은 ROI 필터에서 제공하는 데이터 형에 따라서, DirectShow™의 필터 연결 과정 의해 자동적으로 선택된다.

코드 구현은 *CTransformFilter* 클래스로부터 상속되어 필요한 부분을 추가하는 방법으로 제작되었다. 본 필터는 그림의 X, Y 크기가 변경됨으로 해서 제자리 변환 필터(in place transform filter)를 사용할 수 없었다. 제자리 변환필터는 버퍼 복사를 최소화 함으로써 변환 필터로 인한 성능저하를 최소화하는 설계방법이다.

*IMediaSample* 인터페이스의 *GetTime()* 메소드를 통해서 100n sec 단위의 미디어 플레이 시간을 제공한다. 본 논문에서는 이 메소드를 통해서 ROI 가 시간에 따라 사각형 형태로 움직일 수 있게 하였다.

#### IV. 실험결과

그림 6 에서는 사용자 관심영역에 따른 변환을 하기 전 영상과 후 영상이 나타나 있다. 관심 영역을 중심으로 보다 자세한 영상을 볼 수 있으며, 특히 디스플레이 사이즈가 제한되는 경우에 시각적으로 보다 좋은 영상을 얻을 수 있다. 또한 콘텐츠에 따라서 ROI는 시간에 따라 움직일 수 있다. 제안된 ROI 변환 엔진에서는 MPEG-21 표준에 따라 ROI 정보파일을 읽어서 관심영역을 시간에 따라 움직일 수 있도록 되어 있다. 본 논문에서 제안된 사용자 관심영역 적응변환 엔진을 사용하면 다양한 코덱을 사용하는 다양한 어플리케이션 제작이 가능하다.

디스플레이사이즈가 큰 경우에는 공을 넣는 선수와 골키퍼 및 공을 선명하게 볼 수 있다. 그러나 디스플레이 화면이 작은 경우에는 원 콘텐츠의 단순 축소는 의미상 효과적이지 못하다. 축구공이 보이지 않으며 선수도 명확하지 못하다. 그러나 ROI 적응변환을 적용한 화면에서는 최소한의 영역에서 자세한 영상을 볼 수가 있다. 본 구현에서는 클라이언트에서 ROI정보를 생성 할때, 사용자가 ROI 정보를 입력하고 끝낼 때까지 기다렸다가 이를 서버 측에 보내도록 되어 있다. 좀 더 유연한 적응변환이 되려면 실시간으로 서버와 클라이언트가 ROI정보를 주고 받는 설계가 필요하다. 또한 본 구현에서는 XDI 파일 크기를 고려하여, 초마다 ROI정보가 갱신되도록 하였다. 그러나 픽셀 단위의 ROI 이동은 시작적으로 눈에 거슬린다. 좀더 자연스러운 영상을 위해서는 보간을 이용한 그림 이동방법이 필요할 것이다.

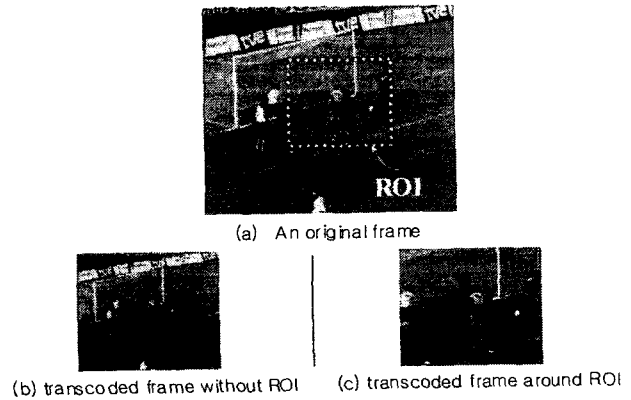


그림 6 관심 영역에 기반한 리소스 적응 변환

#### V. 결론

본 논문에서는 MPEG-21 DIA의 적응변환 엔진 중에 사용자 관심영역 기반의 적응변환 엔진을 구현하고 이에 대한 실험결과를 제시하였다. 모듈화된 컴퍼넌트 구조는 MPEG-21의 다양한 콘텐츠를 사용자의 취향 및 네트워크 환경에 맞게 적응 변환하는데 효과적인 도구이다. 또한, 제안된 적응변환 엔진은 MPEG-21 테스트 베드에서 사용이 매우 유용한 것으로서 다른 적응변환 엔진과 쉽게 연동이 가능하다. 향후에는 여러가지 적응변환 엔진들이 테스트 베드에 집약되어 사용자의 취향에 맞는 다양한 적응변환이 가능해야 할 것이다.

#### 참고문헌

- [1] MPEG document, "MPEG-21 Overview v.5", ISO/IEC JTC1/SC29/WG11 N5231, 62nd MPEG Shanghai, October 2002.
- [2] MPEG document, "MPEG-21 Digital Item Adaptation" ISO/IEC 21000-7 Final Committee Draft, ISO/IEC/JTC1/SC29/WG11 N5845, 65th MPEG Trondheim, July 2003.
- [3] MPEG document, Munchurl Kim, Jeongyeon Lim, Jong-Nam Kim and Kyeongsoo Kim, "Extensions to Presentation Preference DS of MPEG-21 DIA," ISO/IEC JTC1/SC29/WG11 m8968, Shanghai, China, Oct. 2002.
- [4] MPEG document, "MPEG-21 Digital Item Processing Working Draft (WD)", ISO/IEC JTC1/SC29/WG11 N5855, v.2" Trondheim, Norway, Jul. 2003.
- [5] 손유미, 임정연, 김문철, 김종남, 김경수, "사용자의 관심 영역을 고려한 MPEG-21 디지털 아이템 적응 변환," 2003 전자공학회 하계 학술대회, Jul. 2003.
- [6] 신화선, "DirectShow 멀티미디어 프로그래밍", 한빛미디어